Non-parametric regression for patch-based fluorescence microscopy image sequence denoising

Jérôme Boulanger, Charles Kervrann, Jean Salamero, Jean-Baptiste Sibarita, Patrick Bouthemy

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Non-parametric regression for patch-based fluorescence microscopy image sequence denoising

Jérôme Boulanger Charles Kervrann , Jean Salamero ,

Jean-Baptiste Sibarita , and Patrick Bouthemy

N° 6651

September 2008

Thème COG

apport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6651--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Non-parametric regression for patch-based fluorescence microscopy image sequence denoising

Jérôme Boulanger * Charles Kervrann †, Jean Salamero ‡,
Jean-Baptiste Sibarita §, and Patrick Bouthemy ¶

Thème COG — Systèmes cognitifs
Équipe-Projet Vista

Rapport de recherche n° 6651 — September 2008 — 34 pages

Abstract: We present a non-parametric regression method for denoising 3D image sequences acquired in fluorescence microscopy. The proposed method exploits 3D+time information to improve the signal-to-noise ratio of images corrupted by mixed Poisson-Gaussian noise. A variance stabilization transform is first applied to the image-data to introduce independence between the mean and variance. This pre-processing requires the knowledge of parameters related to the acquisition system, also estimated in our approach. In a second step, we propose an original statistical patch-based framework for noise reduction and preservation of space-time discontinuities. In our study, discontinuities are related to small moving spots with high velocity observed in fluorescence video-microscopy. The idea is to minimize an objective nonlocal energy functional involving spatio-temporal image patches. The minimizer has a simple form and is defined as the weighted average of input data taken in spatially-varying neighborhoods. The size of each neighborhood is optimized to improve the performance of the pointwise estimator. The performance of the algorithm which requires no motion estimation, is then demonstrated on both synthetic and real image sequences using qualitative and quantitative criteria.

Key-words: Video-microscopy, fluorescence, image sequence denoising, patch-based approach, Poisson noise, variance stabilization, adaptive estimation

* jerome.boulanger@oeaw.ac.at
† ckervran@irisa.fr
‡ jean.salamero@curie.fr
§ jbs@curie.fr
¶ bouthemy@irisa.fr

Centre de recherche INRIA Rennes – Bretagne Atlantique
IRISA, Campus universitaire de Beaulieu, 35042 Rennes Cedex
Téléphone : +33 2 99 84 71 00 — Télécopie : +33 2 99 84 71 71

Régression non-paramétrique et représentation des images par motifs locaux pour le débruitage de séquences de vidéo-microscopie de fluorescence

Résumé : Nous présentons une méthode de régression non-paramétrique pour le débruitage de séquences d'images volumiques acquises en microscopie de fluorescence. La méthode proposée utilise l'information spatio-temporelle en vue de l'amélioration du rapport signal-à-bruit des images perturbée par un bruit Poisson-Gaussien. Une stabilisation de la variance est préalablement appliquée aux données images afin de rendre la variance du bruit indépendante de l'intensité locale. Dans une deuxième étape, nous proposons un cadre original pour la réduction du niveau de bruit et la préservation des discontinuités spatio-temporelles. Dans notre étude, ces discontinuités sont introduites par de petits objets ayant des vitesse élevées. Notre approche repose sur la minimisation d'une fonctionnelle d'énergie impliquant des motifs contenus dans les séquences d'images. Le minimiseur possède une forme simple et est défini comme une moyenne pondérée de données collectées dans un voisinage spatio-temporel. La taille de chaque voisinage est localement optimisée afin d'améliorer les performances de l'estimateur. Au final, les performances globales de l'algorithme, qui ne requière pas d'estimation du mouvement, sont illustrées à l'aide de séquences d'images synthétiques mais aussi à l'aide de données réelles en utilisant des critères qualitatifs et quantitatifs.

Mots-clés : Vidéo-microscopie, fluorescence, image sequence denoising, patch-based approach, bruit de Poisson, stabilisation de la variance, estimation adaptative

Contents

1 Introduction

Fluorescence video-microscopy is an investigation tool used for dynamics analysis at sub-cellular levels in biology. Combined with fluorescent tags such as genetically engineered fluorescent chimeric proteins (e.g. Green Fluorescence Protein GFP), both confocal microscopy and wide-field microscopy allows 3D live protein imaging. Mainly used to analyze isolated cells, confocal microscopy can also be used *in vivo* if combined with endomicroscopy. Unfortunately, when cell viability needs to be preserved and photobleaching avoided, light exposure time must be limited, resulting in low signal-to-noise ratios.

While improving the signal-to-noise ratio, denoising may allow us to reduce exposure time and therefore to open new opportunities in live cell imaging. Moreover, frame rates can be then increased without increasing radition dose, which could be relevant to capture fast events at sub-cellular levels. Finally, if the objective's point spread function is not affected by denoising, images may still be compatible with a deconvolution process, significantly increasing the performance of restoration algorithms images with low signal-to-noise ratios. As a consequence, performances of object detection and tracking algorithms are improved as well.

Currently, denoising is a widely studied but still an open problem in image processing. Many methods have been described in the literature, and a recent outstanding review can be found in [1]. Methods based on the full knowledge of noise statistics are probably the more efficient. In fluorescence video-microscopy, it is established that the low level of fluorescence is related to a limited number of photons that can be modeled as a Poisson process. Besides, additive electronic noise is usually present even if a cooling system is used on the detector. The resulting images are then assumed to be contaminated by a combination of Poisson and Gaussian noise. Several approaches have been introduced to deal with a signal-dependent noise. In [2], the authors proposed a Maximum Likelihood estimator for Poisson noise removal in very low count situations. The problem is more challenging for Poisson-Gaussian noise and another line of work consists in stabilizing the noise variance using ad-hoc transforms. The more common transform is the so-called Anscombe transform [3] designed for Poisson noise. This transform was further generalized to Poisson-Gaussian noise, with satisfying results if the number of counts is large enough [4] and more recently for "clipped" (under- and over-exposure) raw-data [5]. In the case of very low count situations (≤ 1 photons in average), the more sophisticated Fisz transform allows one to better stabilize Poisson noise [6, 7]. Finally, local estimation of image-dependent noise statistics (assumed to be locally Gaussian) has been investigated, especially in the case of adaptive Wiener filtering [8–10].

Denoising image in temporal sequences is even more complex since there are currently no satisfying methods for processing fluorescence videomicroscopy 3D image sequences contaminated by Poisson-Gaussian noise. Most of them only restore every frame separately without using the temporal redundancy of image series. When temporal coherence is exploited, it is usually recommended to consider

a motion estimation/compensation stage as proposed for video denoising [11–14] and, for instance, for low-dose fluoroscopy image sequence filtering [10]. This is especially true for real-time imaging applications. Thus, Kuznetsov et *al.* recently proposed to use a temporal Kalman-Bucy filter to improve the quality of video-microscopy image sequences [15]. The main difficulty in video-microscopy is to estimate the motion of small and similar objects moving with high velocity in the image sequence. To overcome this problem, sophisticated methods (see [1]) but designed for still images have been adapted to videos. Wavelet shrinkage [16, 17], Wiener filtering [18] or PDE-based methods [19] are typical examples of such methods. Recently, an extension of the non-local means filter [1] also related to the universal denoising (DUDE) algorithm [20] and the entropy-based UINTA filter [21], has been proposed to process image sequences. It assumes that image sequence contains repeated patterns [22]. Noise can then be reduced by averaging data associated to the more similar patches in the image sequence. Patch-based approaches are now very popular in texture synthesis [23], inpainting [24], video completion [25] ; they have also been explored for image restoration [26].

Nevertheless, searching similar examples in the whole image for denoising with the non-local means filter, is untractable in practice in 2D, and unrealistic for video sequences. As a consequence, a variant of this filter has been recently proposed in [27] ; the authors use a pre-classification of pixels in the sequence in order to speed up the denoising procedure. Another improvement introduced in [28] consists in collecting similar patches to build 3D arrays. A unitary transform and a hard-thresholding are then applied to remove noise.

A general modeling framework based on signal and information theory has been proposed by Elad et *al.* for image and video sequence analysis. The authors assume that the images can be approximated by a sparse representation and dictionaries of forms, like DCT coefficients or libraries of patches [29]. The approximation problem is then equivalent to the global minimization (using a a K-SVD algorithm) of an energy functional involving a data term and a penalty term that encodes sparsity [30]. This method is able to produce impressive image denoising results, including video image sequences, but requires intensive minimization procedures and the adjustment of several parameters.

Unlike the previous patch-based approaches [22, 27, 31, 32], we present in this paper an original space-time patch-based adaptive statistical method for 3D+time video-microscopy image sequence restoration. As already mentioned, patch-based methods have been proposed for denoising image sequences, but, to our knowledge, only anisotropic diffusion and wavelet shrinkage have been applied to 2D+time fluorescence video-microscopy [33, 34]. In our approach, we propose first a variance stabilization step to be applied to the data in order to obtain independence between the mean and the variance. Second, we consider spatio-temporal neighborhoods to restore series of 3D images as already proposed in [32] in a discrete setting. Our method is based on the minimization of an energy functional while exploiting image patches. The minimizer of this energy functional has a simple form and corresponds to a weighted average of intensity values taken in spatially (and

temporally) varying neighborhoods. The neighborhood size is adapted on-line to improve the performance (in the sense of L_2 risk) of the pointwise estimator. No learning step or wavelet decomposition is required. Also, no motion estimation is involved as orignally described in [32]. Finally, the designed algorithm comprises only few parameters which are easily calibrated.

The remainder of this paper is organized as follows. In Section 2, we introduce the denoising problem in fluorescence video-microscopy. In Section 3, the main contributions are presented in detail. In Section 4, we demonstrate the performance of the algorithm (controlled by a small number of parameters) on both synthetic and real video-microscopy image sequences.

2 Problem statement

In this section, we present a general framework for image sequence analysis in wide-field or confocal microscopy. Our study is limited to the restoration of artifacts due to noise. We do not consider the issue of correcting the signal distortions due to diffraction (e.g. deconvolution problem). We will later show the compatibility of the proposed method with further deconvolution.

Acquired images correspond to stacks of 10 to 60 slices with an axial resolution (depth) lower than the lateral one. Anisotropy in 3D microscopy can be an issue for 3D wavelet methods, especially for processing stacks with a limited number of slices (boundary effects). The processed images depict tagged proteins appearing as bright particles of size 3 to 10 pixels moving with speeds ranging from 1 to 10 pixels per frame. The small amount of light collected by sensors and thermal agitation in electronic components induce a mixed Poisson-Gaussian noise. Accordingly, we assume the following linear model:

$$Z(x) = g_0 N(x) + \varepsilon(x), \tag{1}$$

where $Z(x)$ is the observation at the space-time location $x \in \mathbb{R}^4$ and g_0 represents the gain of the overall electronic system. The number $N(x)$ of collected photo-electrons is a random variable assumed to follow a Poisson distribution of parameter $\theta(x)$: $p(N(x)) = \frac{\theta(x)^{N(x)} e^{-\theta(x)}}{N(x)!}$. Finally, the dark current is modeled by a Gaussian white noise of variance $\mathrm{Var}[\varepsilon(x)] = \sigma_\varepsilon^2$ and expectation $\mathbb{E}[\varepsilon(x)] = m$. Finally, let's note $f(x) = g_0\theta(x) + m$.

In this paper, we consider the problem of estimating $f(x)$ at each point x from noisy data $Z(x)$ taken in a space-time neighborhood of x. We propose the following contributions:

- First, we assume that the lowest number of detected photo-electrons is large enough (> 30) and we adopt the generalized Anscombe transform to stabilize the noise variance. We need the prior knowledge of the following quantities: g_0, m and σ_ε^2. In Section 3.1, we present a robust data-driven

method to estimate these parameters. Once, the noise variance is stabilized, a more convenient additive Gaussian noise model is considered.

- Second, we minimize an energy functional based on image patches, able to capture local geometries and spatial interactions. Unlike previous methods, we compute a distance between spatio-temporal patches for detecting similarities and redundancies in the 3D+time domain. Furthermore, we show that the fixed-point solution has a simple form: the minimizer involves the weighted average of input data taken in a varying space-time neighborhood. The set of nearby patches can be then interpreted as a variable dictionary of patches which length is related to the size of the neighborhood.

- Third, we adapt locally the number of elements of this dictionary. We progressively increase the number of patches participating to the weighted average by increasing the space-time neighborhood size. The optimal dictionary size is defined as the one that minimizes the local L_2 risk. Following the Lepskii's approach [35, 36], it amounts to balancing the bias and variance of the estimator by adapting locally the size of the neighborhood.

In the next section, we address these three issues in detail.

3 Proposed method

3.1 Stabilization of noise variance

The Anscombe transform is the more commonly-used transform for stabilizing the variance of Poisson noise [3]. In [37], the authors used this transform to denoise confocal images, since the number of photons is large enough. Earlier, Murtargh et *al.* considered a more general Anscombe transform of the form [38]:

$$\mathcal{T}_{GA}(Z(x)) = \frac{2}{g_0}\sqrt{g_0 Z(x) + \frac{3}{8}g_0^2 + \sigma_\varepsilon^2 - g_0 m}. \tag{2}$$

In contrast to the usual parameter-free Anscombe transform, the Generalized Anscombe transform requires the setting (or the estimation) of a small set of parameters, g_0, σ_ε^2 and m, related to the acquisition system. Starck et *al.* proposed an iterative algorithm to estimate the gain g_0 and the dark current parameters from images [39]. Their method stabilizes the variance of the transformed data by testing several parameters according to a dichotomy process. Instead, we have defined an approach based on a linear regression in the 2D-space $(\mathbb{E}[Z(x)], \mathrm{Var}[Z(x)])$. This method has been previoulsy skeched in [40] and we provide here more details. Note that a similar approach has also been recently used in [34]. From (1), we have

$$\begin{cases} \mathbb{E}[Z(x)] &= g_0\theta(x) + m, \\ \mathrm{Var}[Z(x)] &= g_0^2\theta(x) + \sigma_\varepsilon^2. \end{cases} \tag{3}$$

which yields

$$\mathrm{Var}[Z(x)] = g_0 \mathbb{E}[Z(x)] + \sigma_\varepsilon^2 - g_0 m. \tag{4}$$

It follows that a linear regression in the 2D-space $(\mathbb{E}[Z(x)], \mathrm{Var}[Z(x)])$ provides an estimation of the two parameters g_0 and $e_{DC} = \sigma_\varepsilon^2 - g_0 m$. Accordingly, (2) can be written as

$$\mathcal{T}_{GA}(Z(x)) = \frac{2}{g_0}\sqrt{g_0 Z(x) + \frac{3}{8}g_0^2 + e_{DC}}. \tag{5}$$

Now, we robustly estimate the local mean and the local variance. In order to get independent samples and to save computation time, it is crucial to partition the space-time volume into non-overlapping blocks. The size of these blocks results from a compromise between the estimator variances and the number of resulting measure points in $(\widehat{\mathbb{E}[Z(x)]}, \widehat{\mathrm{Var}[Z(x)]})$. For each block, we get a measurement point of coordinates $(\widehat{\mathbb{E}[Z(x)]}, \widehat{\mathrm{Var}[Z(x)]})$. The mean $\widehat{\mathbb{E}[Z(x)]}$ is estimated using a robust M-estimator. The noise variance $\widehat{\mathrm{Var}[Z(x)]}$ is robustly estimated using the "Least Median of Squares" (LMedS) estimator defined as

$$\widehat{\mathrm{Var}[Z(x)]} = 1.4826\, \mathrm{med}_x \left(|r(x) - \mathrm{med}_y\, |r(y)|| \right), \tag{6}$$

where the pseudo-residuals $r(x)$ are computed at each spatial position $x \in \mathbb{R}^4$ (in the 3D+time case) as [41]:

$$r(x) = \frac{1}{\sqrt{l^2 + l}} \Delta Y(x). \tag{7}$$

Here $\Delta Y(x)$ denotes the space-time Laplacian operator involving l surrounding pixels.

Given empirical estimates of the mean and the variance, a simple linear regression is applied to obtain the values of parameters g_0 and e_{DC}. The Generalized Anscombe Transform is then applied to the input data Z to produce new input data Y with Gaussian statistics.

3.2 Patch-based energy functional

Once the noise variance has been stabilized, we consider the following image sequence model:

$$Y(x) = u(x) + \eta(x), \tag{8}$$

where $x \in \Omega$ denotes the pixel location in the space-time volume $\Omega \subset \mathbb{R}^4_+$. The regression function $u(x)$ is the ideal image to be recovered from observations $Y(x) := \mathcal{T}_{GA}(Z(x))$. The errors $\eta(x)$ are now assumed to be independent zero-mean Gaussian variables with variance σ_η^2 theoretically equal to 1. Because of errors and non-stationarities, stabilization is not ideal and the variance needs to be estimated from data Y. The inverse generalized Anscombe transform is applied to the estimate $\widehat{u}(x)$ afterwards to recover $f(x)$, at each spatial position $x \in \Omega$.

To solve the restoration problem, that is to recover the true image function u from noisy data Y, we propose to minimize an original energy functional $J(u, Y)$ able to capture image spatio-temporal redundancy from image patches. Several approaches have been recently proposed in this line of work [42–44], yielding iterative variants of the non-local means filter for 2D still images. In our framework, we propose an energy functional but, unlike [42–44], the determination of involved parameters is data-driven.

Let define $J(u, Y)$ the energy functional as

$$J(u, Y) = \int_{\Omega\times\Omega} \phi\left(d_{\mathbf{Q}}(\mathbf{u}(x), \widetilde{\mathbf{u}}(y))\right) K\left(\frac{\|x-y\|}{h(x)}\right) dy\, dx, \tag{9}$$

where K is an appropriate spatial kernel ($\|\cdot\|$ denotes the usual euclidean norm) with a spatially-varying bandwidth $h(x)$ acting as a space-time neighborhood, $\mathbf{u}(x) := (u(x'))_{\|x-x'\|\leq\rho}$ corresponds to a small image patch (size is parametrized by ρ) for image sequences and $\phi : \mathbb{R}^+ \to \mathbb{R}$ is a differentiable function which can be convex or not. In (9), we consider the usual Mahalanobis distance to compare patches:

$$d_{\mathbf{Q}}(\mathbf{u}(x), \widetilde{\mathbf{u}}(y)) = (\mathbf{u}(x) - \widetilde{\mathbf{u}}(y))^T \mathbf{Q}^{-1}(x, y)(\mathbf{u}(x) - \widetilde{\mathbf{u}}(y)) \tag{10}$$

where $Q(x, y)$ is a covariance function discussed later. In the definition of $\widetilde{\mathbf{u}}(y)$, $Y(y)$ is substituted to $u(x)$ and, accordingly, the minimization of $J(u, Y)$ does not lead to a constant image. More formally, $\widetilde{\mathbf{u}}(y)$ is componentwise defined at point y as :

$$\widetilde{u}(y-z) = \begin{cases} Y(y) & z = 0, \\ u(y-z) & \text{if } 0 < \|y-z\| \leq \rho, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

In what follows, we assume that the patch size is fixed for every point and parametrized by ρ. Finally, we choose K as a cut-off function:

$$K(z) = \begin{cases} 1 & \text{if } |z| \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

and we take $\phi(z) = 1 - e^{-z}$ as also suggested in [45].

$J(u, Y)$ is a non-local energy functional and is improved here by introducing image patches in the definition. Intuitively, minimizing $J(u, Y)$ amounts to estimating an image for which neighboring patches are similar and, at the same time, the estimated value at the central position in the reference patch $\mathbf{u}(x)$ must be as close as possible to the input data $Y(y)$ observed at the central positions in the neighboring patches $\{\mathbf{u}(y)\}$. The non-local and complex interactions in spatially varying neighborhoods are thus taken into account in the framework. According to

the variation calculus method, we have

$$J(u+\delta u,Y)-J(u,Y)= \int_{\Omega\times\Omega}\phi\left[d_{\mathbf{Q}}(\mathbf{u}(x)+\delta\mathbf{u}(x),\widetilde{\mathbf{u}}(y)+\delta\widetilde{\mathbf{u}}(y)) -\phi\Big(d_{\mathbf{Q}}(\mathbf{u}(x),\widetilde{\mathbf{u}}(y))\Big)\right]K\left(\frac{\|x-y\|}{h(x)}\right)dy\,dx, \quad (13)$$

with the abbreviation $\delta\mathbf{u}(x) := (\delta u(x'))_{\|x-x'\|\leq\rho}$. The components of $\delta\widetilde{\mathbf{u}}(y)$ are defined as

$$\delta\widetilde{u}(x-z)=\begin{cases}u(x-z) & \text{if } 0<\|x-z\|\leq\rho,\\ 0 & \text{otherwise.}\end{cases} \quad (14)$$

A first-order Taylor expansion leads to

$$J(u+\delta u,Y)-J(u,Y)\approx 2\int_{\Omega\times\Omega}(\delta\mathbf{u}(x)-\delta\widetilde{\mathbf{u}}(y))^T\mathbf{Q}^{-1}(x,y)(\mathbf{u}(x)-\widetilde{\mathbf{u}}(y)) \phi'\Big(d_{\mathbf{Q}}(\mathbf{u}(x),\widetilde{\mathbf{u}}(y))\Big)K\left(\frac{\|x-y\|}{h(x)}\right)dy\,dx. \quad (15)$$

Since we are only interested in the local variation at point x, we set $\delta u(y)=0,\forall y\neq x$. In addition, if y and x are mutually neighbors, it follows that

$$\frac{J(u+\delta u,Y)-J(u,Y)}{\delta u(x)}\approx\mathbf{Q}^{-1}(x,x) \times\int_{\Omega}(u(x)-Y(y))\,\phi'\Big(d_{\mathbf{Q}}(\mathbf{u}(x),\widetilde{\mathbf{u}}(y))\Big)K\left(\frac{\|x-y\|}{h(x)}\right)dy. \quad (16)$$

If u is a stationary point of $J(u,Y)$, the first-order term vanishes and the fixed-point solution is given by

$$\widehat{u}(x)=\frac{\displaystyle\int_{\Omega}\phi'\Big(d_{\mathbf{Q}}(\mathbf{u}(x),\widetilde{\mathbf{u}}(y))\Big)K\left(\frac{\|x-y\|}{h(x)}\right)Y(y)\,dy}{\displaystyle\int_{\Omega}\phi'\Big(d_{\mathbf{Q}}(\mathbf{u}(x),\widetilde{\mathbf{u}}(y))\Big)K\left(\frac{\|x-y\|}{h(x)}\right)dy}. \quad (17)$$

We can rewrite this expression as the weighted sum of the original data $Y(y)$ as

$$\widehat{u}(x)=\int_{\Omega}w(x,y)Y(y)\,dy \quad (18)$$

where the weights $w(x,y)$ are defined as

$$w(x,y)=\frac{\phi'\Big(d_{\mathbf{Q}}(\mathbf{u}(x),\widetilde{\mathbf{u}}(y))\Big)K\left(\frac{\|x-y\|}{h(x)}\right)}{\displaystyle\int_{\Omega}\phi'\Big(d_{\mathbf{Q}}(\mathbf{u}(x),\widetilde{\mathbf{u}}(z))\Big)K\left(\frac{\|x-z\|}{h(x)}\right)dz}. \quad (19)$$

The solution (17) yields the estimator $\widehat{u}(x)$ for each point x given the original data $Y(y)$. However, the bandwidth $h(x)$ is also unknown. In the next section, we address the issue of estimating the optimal bandwidth controlling the spatial neighborhood. A computational solution will be derived.

3.3 Neighborhood size selection

The estimator (17) is based on the approximation of the central patch by a set of nearby patches. The performance of the estimator is related to the bandwidth h of this neighborhood and can vary at each point of the image sequence according to the image content.

In order to optimally estimate the bandwidth, we analyze the performance of the estimator and consider the usual local L_2 risk defined as

$$\mathcal{R}(\widehat{u}(x), u(x)) = \mathbb{E}\left[(\widehat{u}(x) - u(x))^2\right], \tag{20}$$

where $u(x)$ is the unknown function at point x. The local risk $\mathcal{R}(\widehat{u}(x), u(x))$ is defined at each point x and then differs from usual performance measures that integrate errors on the whole image. A local adaptation of the bandwidth is more appropriate to improve the estimator in the vicinity of discontinuities. In what follows, we aim at estimating the bandwidth $h(x)$ by minimizing $\mathcal{R}(\widehat{u}(x), u(x))$.

First, it is established that (20) can be decomposed into two terms: squared bias $b^2(x)$ and variance $v^2(x)$ as

$$\mathcal{R}(\widehat{u}(x), u(x)) = \underbrace{(\mathbb{E}[\widehat{u}(x) - u(x)])^2}_{b^2(x)} + \underbrace{\mathbb{E}[(\widehat{u}(x) - \mathbb{E}[\widehat{u}(x)])^2]}_{v^2(x)}. \tag{21}$$

A usual form for the considered estimator variance is

$$\widehat{v}_h^2(x) = \sigma_\eta^2 \int_\Omega (w(x,y))^2 \, dy. \tag{22}$$

The bias term $b(x)$ depends on the unknown function $u(x)$ and is thus unobservable. However, assuming minimal properties about the unknown function u, we can propose an upper bound for the squared bias term. First, we assume that the function u is continuous Lipschitz (in $\mathbb{R}^d$), that is

$$\exists\, C_1 \in \mathbb{R}^{+*} \ : \ |u(x) - u(y)| < C_1 \|x - y\|. \tag{23}$$

and assume

$$\exists\, h(x) \in \mathbb{R}^{+*} \ : \ w(x,y) = 0 \text{ if } \|x - y\| > h(x). \tag{24}$$

From (23) and (24), it comes

$$\begin{aligned} |b_h(x)| &= \left| \int_\Omega w(x,y)\mathbb{E}[Y(y)]\,dy - u(x) \right| \\ &\leq \int_\Omega w(x,y)\,|u(y)-u(x)|\,dy \\ &\leq C_1 \int_\Omega w(x,y)\,|y-x|\,dy \\ &\leq C_1 h(x). \end{aligned} \tag{25}$$

Other more accurate upper bounds can be obtained for this term [46, 47]. More generally, we can consider the general upper bound of the form $b_h^2(x) \leq C_1^2 h^2(x)$ [46–48]. Similarly, the variance is usually bounded as: $v_h^2(x) \leq \sigma_\eta^2 h^{-d}(x)/C_2$ with C_2 a strictly positive constant. From (21), it follows that

$$\mathcal{R}(\widehat{u}(x), u(x)) \leq C_1^2 h^2(x) + \frac{\sigma_\eta^2}{C_2} h^{-d}(x). \tag{26}$$

Our goal is to determine $h(x)$ so that the bound for the risk is minimized. The optimal value of $h(x)$ can be easily obtained:

$$h^*(x) = \left(\frac{d\sigma_\eta^2}{2C_2C_1^2} \right)^{\frac{1}{d+2}}. \tag{27}$$

Unfortunately, this expression of $h^*(x)$ still depends on unknown constants, C_1 and C_2. From (27), the expressions of the bias and variance for the ideal value $h^*(x)$ are

$$b_{h^*}^2(x) \approx C_1^2 \left(\frac{d\sigma_\eta^2}{2C_2C_1^2} \right)^{\frac{2}{d+2}}, \tag{28}$$

$$v_{h^*}^2(x) \approx \frac{\sigma_\eta^2}{C_2} \left(\frac{d\sigma_\eta^2}{2C_2C_1^2} \right)^{-\frac{d}{d+2}}. \tag{29}$$

Finally, we point out that the ratio of the squared bias and the variance has a simple expression for the ideal value $h^*(x)$:

$$\frac{b_{h^*}^2(x)}{v_{h^*}^2(x)} = \frac{d}{2} \triangleq \gamma^2 \tag{30}$$

and does not depend on C_1 (image regularity) and further is image-dependent [48]. Following the Lepskii's principle [35], we exploit this property to minimize the L_2 risk $\mathcal{R}(\widehat{u}(x), u(x))$. The idea is to design a sequence of increasing bandwidth : $\mathcal{H} = \{h_n(x), n \in [0, N[:\ h_{n-1}(x) \leq h_n(x)\}$. Assuming that the variance $v_n^2(x)$ is a decreasing function of n, the number of samples taken into account is progressively increased to reduce the estimator variance while controlling the

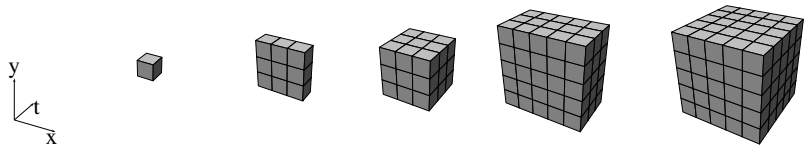


Figure 1: Example of growing space-time neighborhoods in the case of a 2D image sequence. In this representation, each cube represents a pixel in the space-time domain. The proposed mechanism allows us to consider space and time directions as independent components.

estimator bias. Formally, the so-called "bias-variance trade-off" corresponds to the following inequality :

$$h^*(x) = \sup_{h_n(x)\in\mathcal{H}} \{|b_{h_n}(x)| \leq \gamma \upsilon_{h_n}(x)\}. \tag{31}$$

This stepwise procedure will allow us to select, among a predefined set of bandwidths $\{h_n(x), n \in [0, N[\}$, the bandwidth that minimizes the local quadratic risk. Since $b_{h_n}(x)$ is unknown, we consider instead a weaker "oracle" to detect the optimal window for smoothing:

$$h^*(x) = \sup_{h_n(x)\in\mathcal{H}} \{n' < n \ : \ |\widehat{u}_n(x) - \widehat{u}_{n'}(x)| \leq \varrho \upsilon_{n'}(x)\}. \tag{32}$$

where ϱ is a positive constant. We refer the reader to the appendix for the proof. We can notice that this expression involves the comparison between the current estimate and all the previous estimates. Finally, this simple stopping rule allows us to control the risk of the estimator by selecting the optimal bandwidth.

The design of a sequence of increasing bandwidths is now required for estimation. However, when processing an image sequence, the relationship between the temporal and spatial dimensions is related to the object size and movement, which are both unknown. Accordingly, the space and time extents of neighborhoods should be considered independently. For this reason, we decide to increase the size of the neighborhood in an alternate way in space and time, using two distinct bandwidths. We note respectively h^s and h^t the spatial and temporal bandwidths for space and time, and Fig. 1 illustrates the increase of the bandwidths. It is worth noting that, unlike [49], the sequence of bandwidths is not known in advance since we consider two parameters h^s and h^t ; the growing process can be different from one point to another. In our experiments, we use a dyadic scale in space and a linear scale in time.

Finally, given the proposed growth mechanism, the estimator is computed as follows. In (19), the weights depend on the unknown function u at position y. At the initialization, we set $\widehat{u}_0(x) = Y(x)$ since only the noisy data $Y(y)$ are

available. For the next iterations, $n \geq 1$, a new estimate of u can be computed from the previous estimate $\widehat{u}_{n-1}(x)$. Accordingly, the weights are now defined as

$$w(x,y) \approx \frac{\phi'\left(d_{\widehat{\mathbf{Q}}_{n-1}}(\widehat{\mathbf{u}}_{n-1}(x),\widehat{\mathbf{u}}_{n-1}(y))\right)}{\displaystyle\int_\Omega \phi'\left(d_{\widehat{\mathbf{Q}}_{n-1}}(\widehat{\mathbf{u}}_{n-1}(x),\widehat{\mathbf{u}}_{n-1}(y))\right) K\left(\frac{\|x-y\|}{h_n(x)}\right) dy} \tag{33}$$

and the estimator is defined as

$$\widehat{u}_n(x) = \int_\Omega w(x,y) Y(y)\, dy. \tag{34}$$

Besides, the covariance $\widehat{\mathbf{Q}}_{n-1}(x,y)$ is a diagonal matrix which components $\widehat{\mathbf{Q}}_{n-1}(x,y)$ are defined as

$$\widehat{\mathbf{Q}}_{n-1}(x,y) = \lambda^2 \widehat{v}^2_{n-1}(x)\, \delta(x,y)$$

where $\delta(x,y) = 1$ if $x = y$ and 0 otherwise (Kronecker symbol). Once normalized, the distance $d_{\widehat{\mathbf{Q}}_{n-1}}(\widehat{\mathbf{u}}_{n-1}(x),\widehat{\mathbf{u}}_{n-1}(y))$ follows a $\chi^2_{1-\alpha}$ distribution with $|B_\rho(x)| - 1$ degrees of freedom and level $1-\alpha$ where $|B_\rho(x)|$ denotes the patch size. The parameter λ is a α-quantile of the χ^2 distribution. The iteration is stopped at point x if the rule (32) is satisfied. In other words, the estimated spatial or temporal window is defined as

$$\widehat{h}^r(x) = \sup_{h^r_n(x)\in\mathcal{H}^r} \{n' < n \,:\, |\widehat{u}_n(x) - \widehat{u}_{n'}(x)| \leq \varrho \upsilon_{n'}(x)\}, \tag{35}$$

where ϱ is chosen in the range $[2,3]$ and the superscript r denotes s or t.

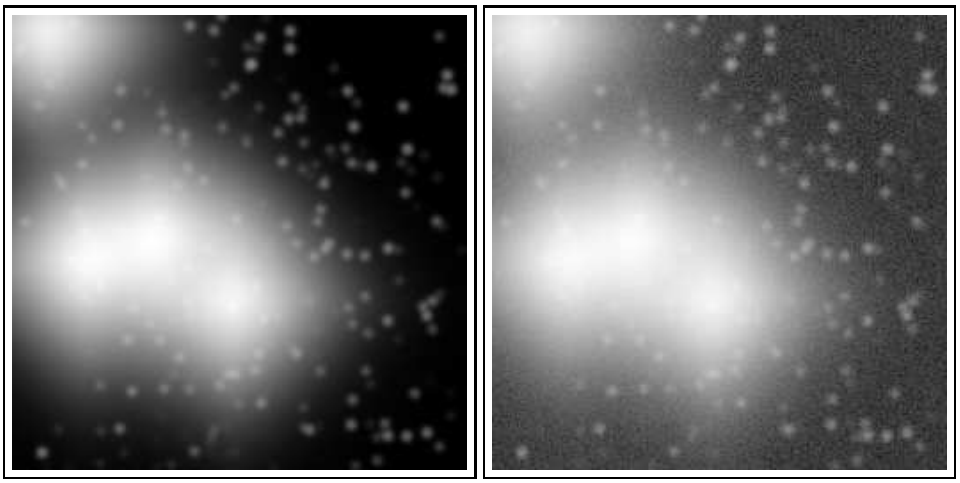
We have now completely described the proposed image sequence restoration method. We can notice that the only free parameter is the patch size, related to the scale of textures and patterns in image sequences. In video-microscopy, the objects are small spots of size ranging from 2 pixels to 4 pixels. Therefore, for the sake of simplicity, $3 \times 3 \times 3$ and $5 \times 5 \times 5$ cubic patches will be considered in our experiments.

4 Experiments

4.1 Synthetic image sequence

In order to test the proposed method, we have generated synthetic image sequences representing moving tagged vesicles. Using this procedure, we aim at analyzing the influence of the generalized Anscombe transform on the final result and to demonstrate that the proposed space-time adaptive method outperforms the corresponding denoising methods used for still images described in [49].

First, we have created a synthetic image sequence showing moving objects superimposed on a static background. The true image sequence is then composed of 50-16bits 3D volumes of $256 \times 256 \times 10$ voxels. The background is generated



(a) (b)

Figure 2: Volume of $256 \times 256 \times 10$ voxels exacted from a simulate image sequence (slice $\#5$ and time $t = 25$), (a) ground truth (b) noisy image sequence (logarithmic scale).

using two or three Gaussian profile of radius 20 pixels with random locations. The background is an essential component of the photometric dynamic of images and thus can alter the stabilization transform process. Typically, the background may be associated to autofluoresence within the cell as well as the non specific accumulation of fluorescent tags on organelles. The flux of photo-electrons related to this component ranges from 10 to 2000 photo-electrons per pixel. Then, 256 spots are drawn as 3D Gaussian functions of radius 2 pixels and of intensity 200 photo-electrons. Their movements are assumed to be described by a Gaussian random walk of standard deviation of 3 pixels. A Poisson noise is generated from this image of flux. Then a gain $g_0 = 0.4$ is applied and finally the dark current is simulated with a Gaussian noise of mean $m = 100$ and a standard deviation $\sigma_\varepsilon^2 = 4$. These values have been obtained by statistical analysis of photometric properties observed in real image sequences. The synthetic image sequence is composed of small spots with intensities of 70 gray levels above the background level, and of 4 large blobs with a maximal intensity of about 900. The slice $\#5$ extracted from a volume at time $t = 25$ of the simulated noise free ground truth and the corresponding noisy slice are shown respectively in Fig. 2(a) and (b).

A scatter plot of the estimated mean and noise variance is shown in Fig. 3(a). The regression line for the first image of the sequence is estimated as $\widehat{\text{Var}[Y(x)]} = 0.407\ \widehat{\mathbb{E}[Y(x)]} - 24.44$, while the theoretical equation is $\text{Var}[Y(x)] = 0.36\ \mathbb{E}[Y(x)] - 24.0$. We can analyze the accuracy of the estimation by considering the next volumes of the sequence. We found that the mean of g_0 is 0.408 and the standard deviation is $6.79 \cdot 10^{-3}$. For the parameter e_{DC}, the mean is -24.31 and the standard deviation 0.879. Accordingly, we can conclude that, for this simulation, the

L_p **norm**	
L_∞	$\sup_{x\in\Omega} \|f(x) - \widehat{v}(x)\|$
L_1	$\int_{x\in\Omega} \|f(x) - \widehat{v}(x)\|\, dx$
L_2	$\int_{x\in\Omega} \|f(x) - \widehat{v}(x)\|^2 dx$

Table 1: Definitions of L_p norms used for evaluation.

parameters of the generalized Anscombe transform has been well estimated. In addition, Fig. 3(b) shows that the variance of the noise has been well stabilized: the noise variance is now 1.001. The width of the cloud of points is related to the errors to the estimation of the mean and noise variance. However, the global trend is well estimated and the noise variance is reliably stabilized.

Our approach is thus quite effective at stabilizing the noise variance in the case of a mixed Poisson-Gaussian noise. It is fully automatic and fast (the computation time of an unoptimized C++ implementation is about 100ms for a single 2D 512×512 image on a 1.8Ghz PC).

To demonstrate the performance of both the variance stabilization procedure and the 3D+time denoising procedure, we consider three experiments: In experiments A and B, we assume respectively a Poisson-Gaussian noise model and a Gaussian noise model. In experiment C, we assume a Poisson-Gaussian noise model but each volume of the sequence is denoised independently. In these three experiments, we used $5\times5\times5$ patches and the algorithm parameters are unchanged.

As reconstruction error, we measured the L_∞, L_1 and L_2 norms (see Table 1) between the original sequence f and the reconstructed image sequence $\widehat{f}$ to compare the different methods and noise models. The results are reported in Table 4.1 and Fig. 4. We can first notice that the L_∞ norm has a high standard deviation. Accordingly, experiments A and C equally supply a better result than C based on the L_∞ norm. This criterion clearly demonstrates that the proposed adaptive modeling is relevant. The use of L_1 and L_2 norms also indicates that the proposed algorithm corresponding to experiment A outperforms the two other techniques (respectively B and C).

Finally, with the peak value of the spots, we consider the peak signal-to-noise ratio PSNR $= 10\log_{10}(\text{Var}[f]/\|\hat{f}(x) - f(x)\|^2)$. We obtain the following results : $PSNR = 24.0$dB, $PSNR = 33.04$dB, $PSNR = 31.06$dB and $PSNR = 32.55$dB respectively for denoised image sequences corresponding to experiments A, B and C.

Besides, the visualization of the result of the restored sequence volume by volume makes clearly appear a flickering artifact due to the lack of coherence between consecutive volumes. In Fig. 6 we can notice the differences between experiments A and B. Flickering artifacts are visible in Fig. 6(b) corresponding to experiment B while in Fig. 6(a) the temporal coherence is reinforced. We can also remark

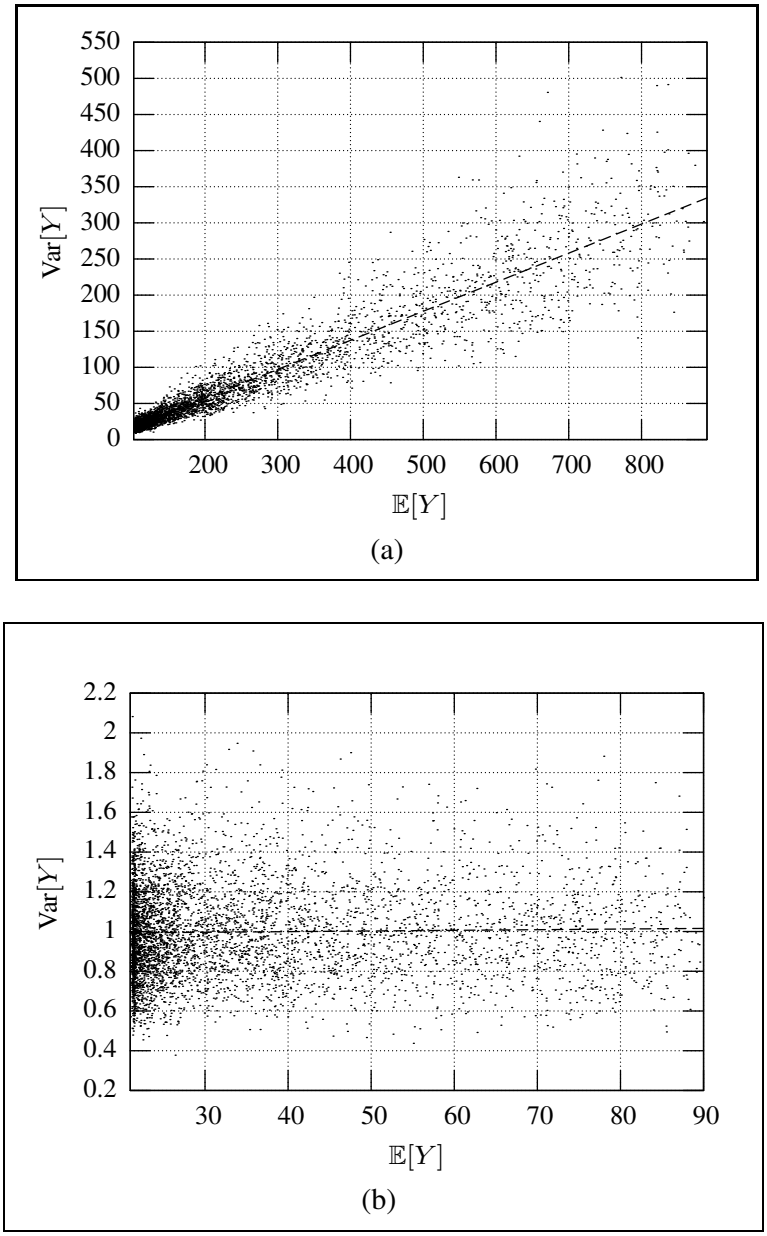


Figure 3: Noise variance stabilization for a synthetic image sequence. Robust estimation of the local mean $\mathbb{E}[Y(x)]$ and noise variance $\text{Var}[Y(x)]$ (a) before stabilization and (b) after stabilization. Each dot (there is a total of 5408 dots) corresponds to a couple $(\widehat{\mathbb{E}[Y(x)]}, \widehat{\text{Var}[Y(x)]})$ estimated on $5 \times 5 \times 5$ non-overlapping blocks. The dashed line represents the fit of the theoretical model $\text{Var}[Y(x)] = g_0\mathbb{E}[Y(x)] + e_{DC}$. After stabilization of the variance, the estimated parameters show no more dependency between the noise variance and the intensity.

Sequences	L_∞		L_1		L_2		$\mathbf{t_e}$
	mean	**std**	**mean**	**std**	**mean**	**std**	
Noisy	62.67	4.21	4.39	$6 \cdot 10^{-3}$	35.0	$12 \cdot 10^{-3}$	
A	38.35	2.87	1.56	$16 \cdot 10^{-3}$	2.94	$28 \cdot 10^{-3}$	65 min
B	53.10	5.83	1.96	$17 \cdot 10^{-3}$	3.78	$25 \cdot 10^{-3}$	55 min
C	37.98	2.44	1.65	$14 \cdot 10^{-3}$	3.01	$24 \cdot 10^{-3}$	28 min

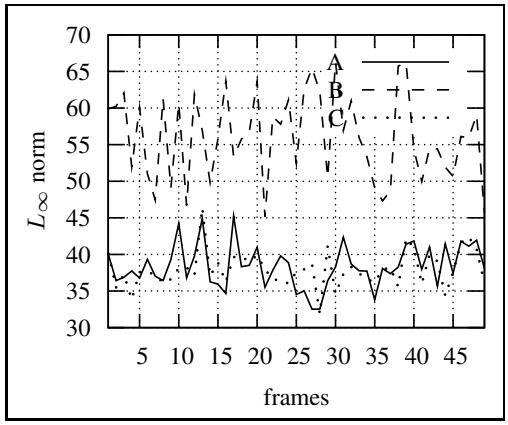
Table 2: Influence of the variance stabilization transform and of the adjacent temporal volumes on the error. Three norms are used to measure the performance of the denoising method. The mean and standard deviation with respect to time are reported. The computation times t_e for each experiment is also given for the noisy sequence; 3D+time - Gaussian and Poisson noise (A) ; 3D+time - Gaussian noise (B) ; 3D - Poisson and Gaussian noise (C).

that temporal discontinuities are well preserved. As expected, these experiments confirm that considering the whole image sequence provides better results than processing the sequence, volume by volume.

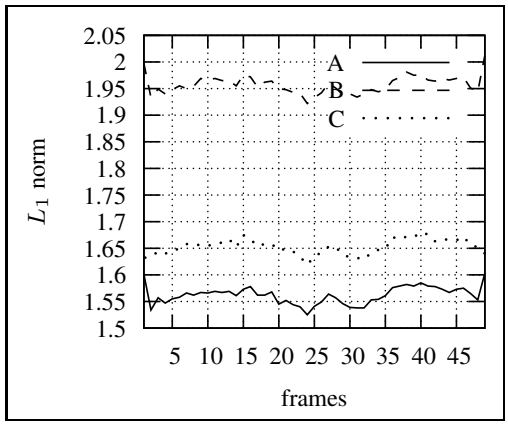
4.2 Spatial denoising of real samples using various exposure times

In this section, we consider several spinning disk acquisitions of the same fixed HeLa cell expressing a GFP tagged Rab6 proteins. For these experiments, the exposure time varies from 30ms to 500ms. Acquired 3D stacks are denoised using a 5×5 median filter and using the proposed method. In this case only 3D information and the Poisson and Gaussian noise modeling is considered since cell are fixed. Results, shown in Fig. 7 reveal that median filtering is not able to both preserve discontinuities and reduce the noise level.

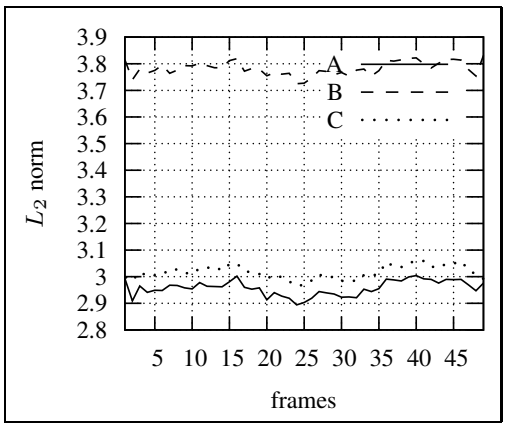
In order to better quantify the potential gain on this real data set, we propose to align the histogram of each 3D image from Fig. 7 onto the histogram of the original raw image obtained with a 500ms exposure time. The alignment is performed by assuming that the relationship between the intensity of an image with the given reference is a linear model and by minimizing the squared errors using a linear regression. This operation does not compensate possible motion between images. However, in this experiment, excepted for $t = 50$ms, the images are aligned. Moreover, motion compensation would imply interpolating of noisy data and could therefore introduce potential artifacts. Figure 8 shows that the L_2 error distance between the reference and the denoised images is lower if the proposed method is applied. For example, the image quality of a 50ms exposure time image processed using the proposed method is approximately the same than for a 200ms exposure time raw acquisition and similar to a 100ms exposure time processed with the median filter. Nevertheless, we should point out that these performances highly



(a)

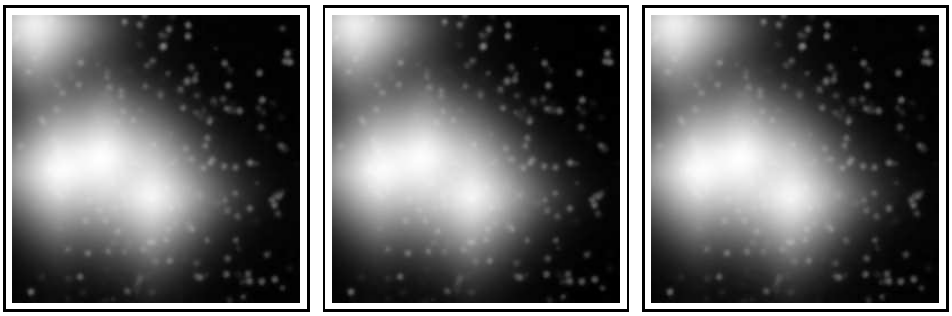


(b)



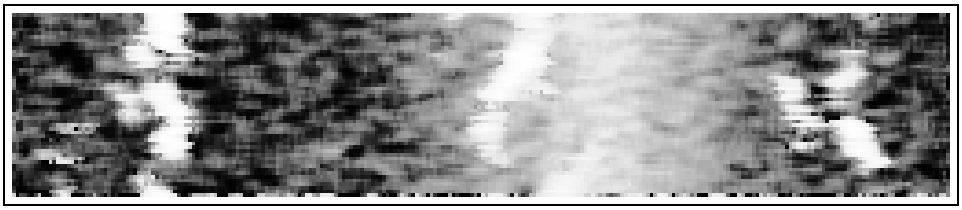
(c)

Figure 4: Influence of the variance stabilization transform and the adjacent temporal volumes on the signal-to-noise ratios. (See text)

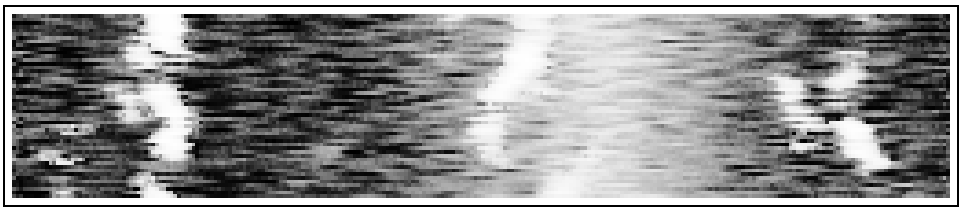


(a) (b) (c)

Figure 5: XY slices $\#5$ at time $t = 25$ of the denoised synthetic image sequence corresponding to experiments A , B and C, respectively in (a), (b) and (c) (logarithmic scale).



(a) 3D+t denoising



(b) 3D denoising

Figure 6: YT slice $\#5$ at $x = 250$ of the denoised synthetic image sequence corresponding to experiments A and C, respectively in (a) and (b) after histogram equalization. More flickering effects are visible when the volumes are independently processed.

depend on the image content. Finally, as previously shown in Section 4.2, the performance of the proposed method would be improved using temporal information.

4.3 Real 3D+time image sequence

We propose now to test the proposed denoising method on a real 3D+time image sequence composed of 50 volumes of $696 \times 520 \times 6$ voxels. The slice $\#3$ extracted at time $t = 20$ is displayed in Fig. 10(a). This sequence has been acquired using a "fast" 4D wide-field microscope. The biological sample is a chimeric protein construct between GFP and Rab6A (GFP-RAB6A) a member of the Rab-GTPases proteins reversibly bounded to specific membranes within the living cell. At the

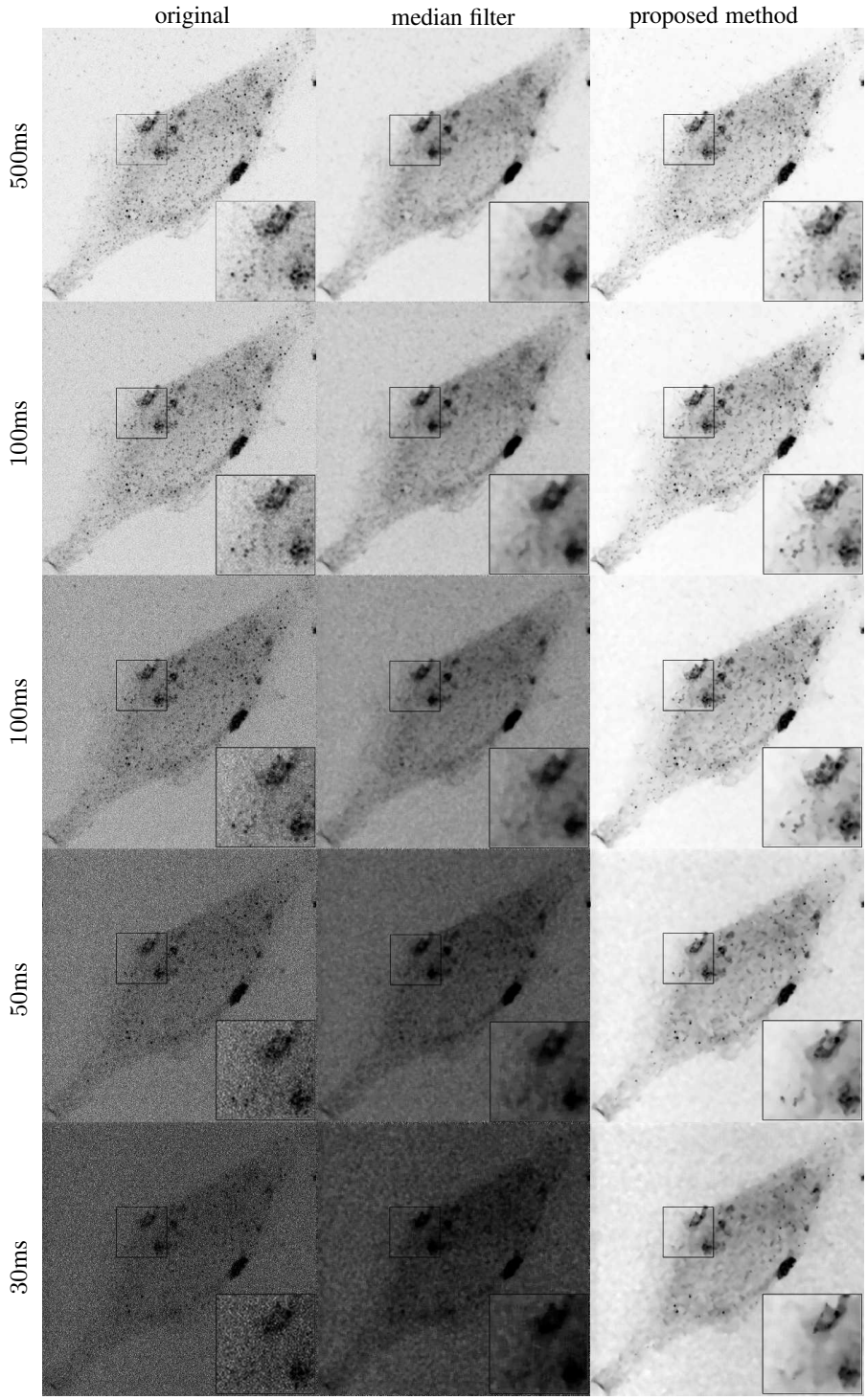


Figure 7: Experiment on a fixed HeLa cell tagged with GFP-Rab6 acquired in spinning disk microscopy. The first column contains 2D slice of original 3D images taken with exposure times ranging from 30ms to 500ms. The second and third columns represent the denoising results obtained respectively with a median filter and the proposed method.

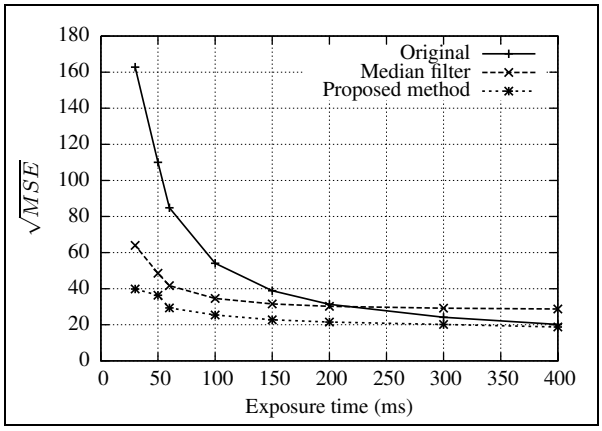


Figure 8: Square root of the mean squared error is plotted against exposure times showing the improvement of the filtering in the case of a fixed sample (see Fig. 7).

steady state, this protein is associated to the Golgi apparatus as well as to rapidly moving transport intermediates and present in the cytosol. Cellular dynamics of Rab6A is influenced by at least three distinct phenomena: i) lateral diffusion dictated by lipid movement within a continuum of membranes ; ii) continuous exchange between cytosolic and membrane bound pools ; iii) directional motion on membrane transport intermediates. In the sequence, the Rab6A proteins appear as bright spots when associated to small moving vesicles inside the living cell. The large bright stable structure corresponds to the Golgi apparatus and the background of the cell reveals its presence into the cytosol.

The estimation of the parameters of the generalized Anscombe transform is illustrated in Fig. 9. The regression lines has been estimated to $\widehat{\mathrm{Var}[Y(x)]} = 0.359\ \widehat{\mathbb{E}[Y(x)]} - 23.36$. As shown in Fig. 9(b), once stabilized, the noise variance is 1.008. The results obtained with our denoising method (using $5 \times 5 \times 5$ patches) are reported in Fig. 10(b). Again, we can notice that the noise has been strongly reduced and that fine details like fluorescent particles are well preserved. The computation time for the whole volume sequence is about 80min using a standard C++ implementation. Experiments on numerous volume sequences confirm the ability of the proposed method to preserve space-time discontinuities.

4.4 Denoising and deconvolution

Wide-field deconvolution microscopy has been widely used this last twenty years in cell biology [50, 51] as a regular tool for monitoring the living cell activity at high spatial and temporal resolution. Compared to confocal like microscopy, it has the advantage to be faster, because of the wide-field illumination, and more efficient thanks to the absence of pinhole to reject photons and the highest quantum efficiency of detectors. Out-of-focus information is used and computationally reassigned to its original place, therefore increasing contrast and signal-to-noise ratio.

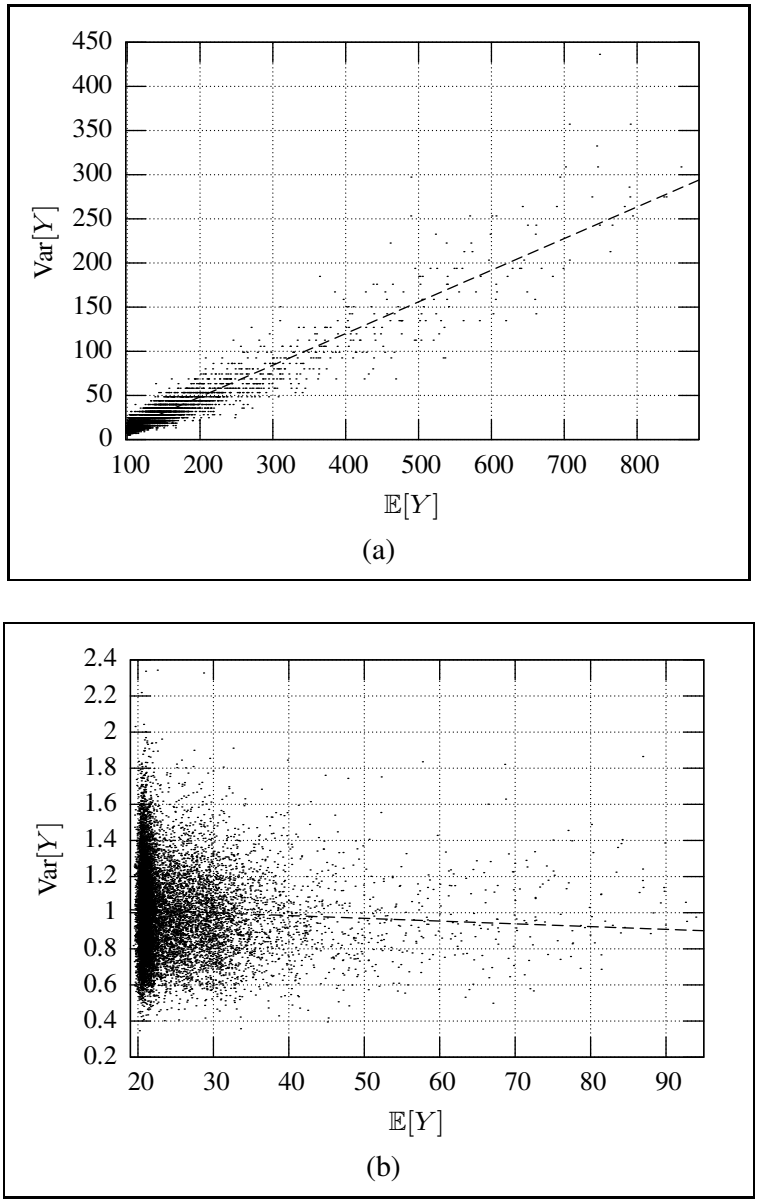
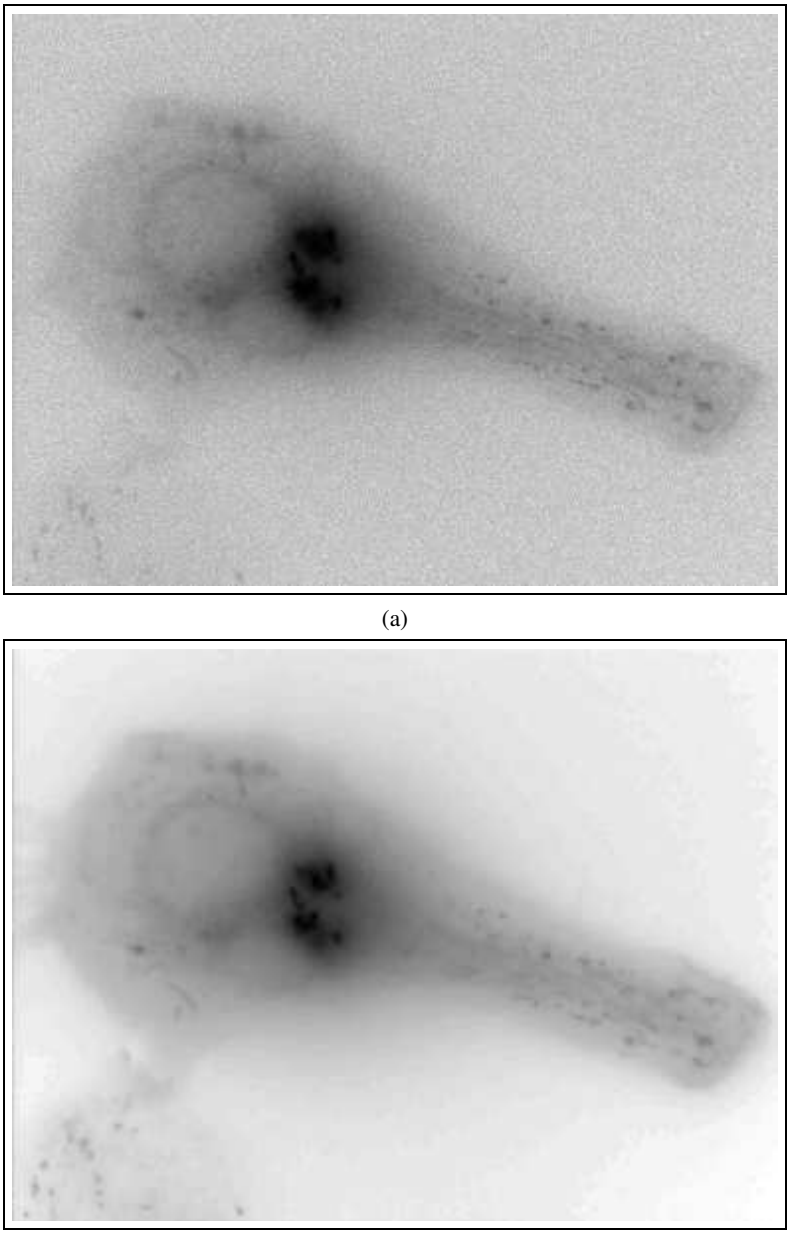


Figure 9: Noise variance stabilization for the real image sequence shown in Fig. 4.3. Estimation of the local mean $\mathbb{E}[Y(x)]$ and local variance $\mathrm{Var}[Y(x)]$ (a) before stabilization and (b) after stabilization.

It is known that the two main limitations of photonic microscopy are i) spatial resolution due to diffraction limit of optics and ii) the number of photons reaching the detector to statistically form the diffraction limited image. In modern living cell microscopy, the number of photons is decreased as much as possible in order to reduce the radiation dose on the sample to keep the cell alive and to increase the acquisition frame rate. The strongest limitation quickly resides in the limited



(a)

(b)

Figure 10: Denoising of a wide-field microscopy image sequence of 50 volumes of size $696 \times 520 \times 6$ voxels. The slice $\#3$ of the original volume at time $t = 20$ is displayed in (a) and the corresponding denoised volume is shown in (b) (logarithmic scale).

number of emitted photons reaching the detector to form an image that can later be described. In addition, deconvolution algorithm efficiency is sensitive to the image signal-to-noise ratio (SNR). The smaller the SNR is the less the algorithms are ca-

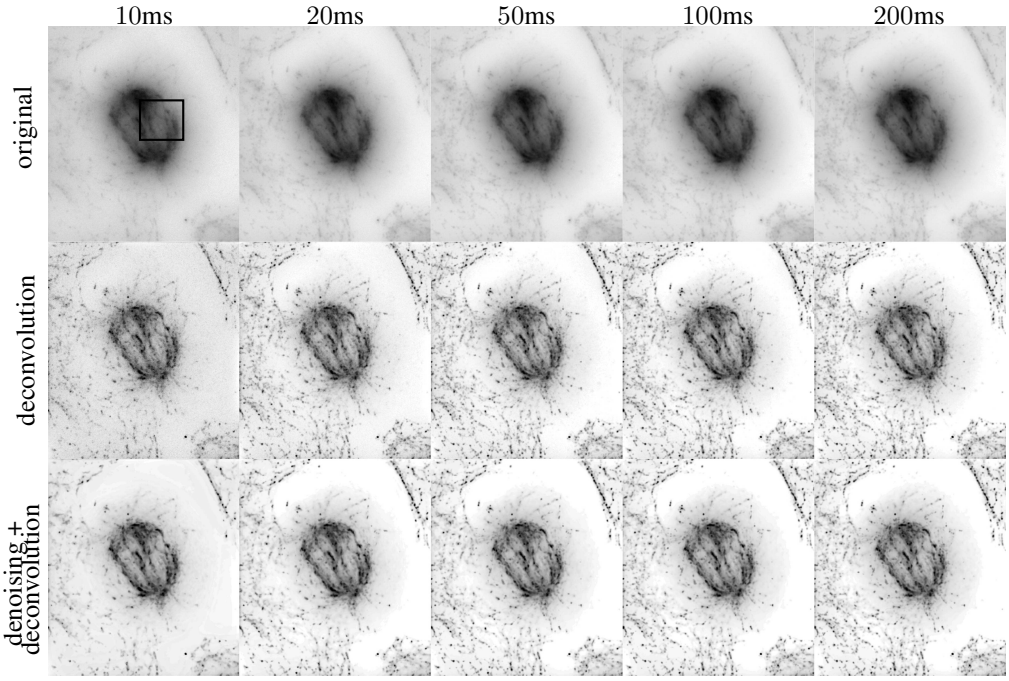


Figure 11: A fixed Hela cell is acquired with five increasing exposure times. The first row contains the maximum intensity projection along z direction of the $200 \times 200 \times 36$ original images. The two last rows correspond respectively to results obtained with the Gold-Meinel deconvolution algorithm [52] and its combination with the proposed patch-based denoising.

pable to restore the relevant signal from the noise, up to not being able to make the difference between noise and signal, resulting in artifacts.

In this section, we propose to combine the proposed denoising approach with an iterative constrained Gold-Meinel deconvolution method [52] using a fixed biological sample. In the same fashion than in Section 4.2, we propose to compare stacks acquired with several exposure times ranging from 10ms to 100ms to a reference image acquired with an exposure time of 200ms. Figure 11 shows the maximum intensity projection of the results. The intensity of original image shown in the first raw ranges from $96 - 260$ gray levels for the image acquired at 10ms of exposure time to $124 - 3315$ gray levels for the image acquired at 200ms of exposure time. Figure 12 shows a zoomed area of an optical section and intensity profiles along a microtubule. It illustrates that fine details are better preserved and that the noise level is strongly reduced. Finally, mean squared errors, computed on normalized images and displayed in Fig. 13, confirm that the deconvolution is improved if the denoising is applied beforehand.

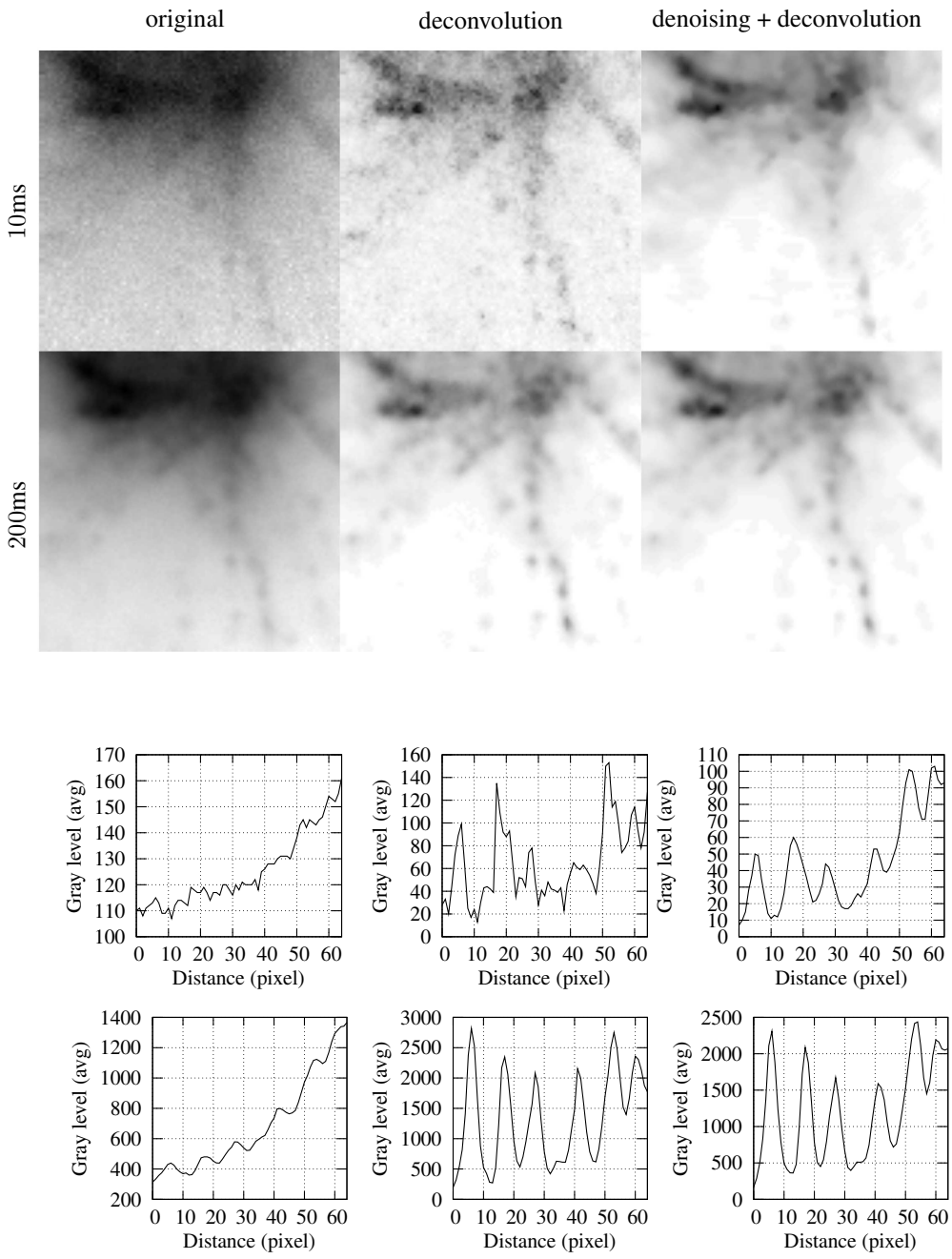


Figure 12: Zoom of a single optical section extracted from data shown in Fig. 11 corresponding to the two extreme exposure times. The colums correspond respectively to the maximum intensity of the raw image, the results obtained with the Gold-Meinel deconvolution algorithm [52] and its combination with the proposed patch-based denoising. Plots show intensity profiles along a single microtubule for each image.

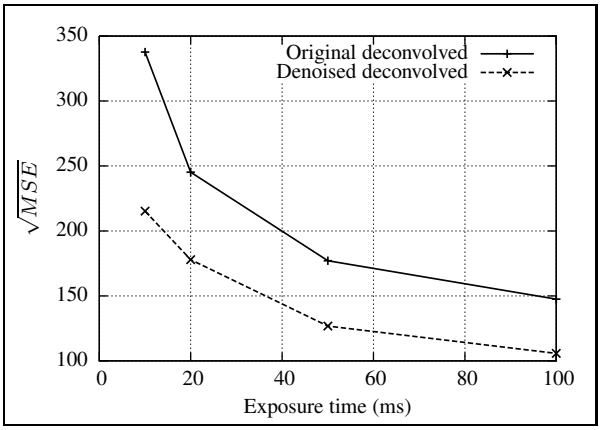


Figure 13: The square root of the mean squared errors is plotted against exposure times in the case of a fixed sample shown in Fig. 11. In one case the Gold-Meinel deconvolution algorithm is applied directly to the original data while on the other case the proposed denoising method is applied.

5 Conclusion

In this paper, we have first tackled the issue of modeling a 3D+time video-microscopy image sequence. We have then proposed to use the generalized Anscombe transform to stabilize the variance of the Poisson and Gaussian noise and we have designed a fast and automatic method to estimate the involved parameters. We have introduced a patch-based functional and we have shown that the fixed-point solution yields an estimator involving image patches taken in a spatially-varying neighborhood. The analysis of the bias-variance of this estimator enabled to properly select, for each point of the space-time domain, the optimal bandwidth within a sequence of increasing bandwidths. Spatial and temporal dimensions are adequately handled. The overall method involves a limited number of parameters so that we do not have to tune them in practice. We have demonstrated that the proposed method outperforms methods based on a Gaussian noise model or processing the sequence volume by volume. Moreover, experiments on real image sequences show that space-time discontinuities are well preserved without optical flow computing. Finally, we have used the capability of the proposed algorithm to efficiently denoise 3D images in order to use it as a preprocessing step prior deconvolution. We have illustrated the efficiency of such a combination to restore low signal-to-noise ratio images. This opens interesting perspectives for monitoring biological samples at high temporal and spatial resolution, without increasing the radiation dose. To conclude, we point out that the proposed method is not restricted to video-microscopy, but could deal with other 2D+time as well as 3D+time noisy image modalities, providing that an appropriate noise modeling is adopted. In this respect, this "breaking sensitivity barrier" approach advantageously complete "breaking resolution barrier" new optics [53].

References

[1] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *SIAM J. Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[2] R. M. Willett and R. D. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Trans. on Medical Imaging*, vol. 22, no. 3, pp. 332–350, Mar. 2003.

[3] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, pp. 246–254, 1948.

[4] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis, the Multiscale Approach*. Cambridge University Press, 2000.

[5] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. on Image Processing*, 2008.

[6] P. Fryzlewicz and G. P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *Journal of Computational and Graphical Statistics*, vol. 13, no. 3, pp. 621–638, Sept. 2004.

[7] B. Zhang, M. Fadili, and J.-L. Starck, "Multi-scale variance stabilizing transform for multi-dimensional Poisson count image denoising," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP'06*, Toulouse, France, May 2006.

[8] J. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Compututer Graphics and Image Processing*, vol. 17, no. 1, pp. 24–32, Sept. 1981.

[9] D. T. Kuan, A. A. Sawchuk, T. V. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 165–177, Feb. 1985.

[10] C. L. Chan, A. K. Katsagellos, and A. V. Sahakian, "Image sequnence filtering in quantum-limited noise with application to low-dose fluoroscopy," *IEEE Trans. on Medical Imaging*, vol. 12, no. 3, pp. 610–621, Sept. 1993.

[11] J. C. Brailean, R. P. Kleihorst, S. Efstratiadis, A. K. Katsaggelos, and R. L. Lagendijk, "Noise reduction filters for dynamic image sequences: A review," *Proc. of the IEEE*, vol. 83, no. 9, pp. 1272–1291, Sept. 1995.

[12] R. Dugad and N. Ahuja, "Video denoising by combining Kalman and Wiener estimates," in *Proc. of IEEE Int. Conf. on Image Processing, ICIP'1999*, vol. 4, Kobe, Japan, Oct. 1999, pp. 152–156.

[13] H.-Y. Cheong, A. Tourapis, J. Llach, and J. Boyce, "Adaptive spatio-temporal filtering for video-denoising," in *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2004*, vol. 2, Singapore, Oct. 2004, pp. 965–968.

[14] V. Zlokolica and W. Philips, "Motion and detail adaptive denoising in video," *Proc. of SPIE – Image Processing: Algorithms and Systems III*, vol. 5298, pp. 403–412, May 2004.

[15] A. Kuznetsov, V. P. Bindokas, J. D. Marks, and L. H. Philipson, "FRET-based voltage probes for confocal imaging: membrane potential oscillations throughout pancreatic islets," *American Journal Physiology - Cell Physiology*, vol. 289, no. 1, pp. 224–229, Mar. 2005.

[16] N. Rajpoot, Z. Yao, and R. Wilson, "Adaptive wavelet restoration of noisy video sequences," in *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2004*, Singapore, 2004, pp. 957–960.

[17] F. Shi and I. W. Selesnick, "Video denoising using oriented complex wavelet transforms," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP'04*, vol. 2, Montreal, Canada, May 2004, pp. 949–952.

[18] F. Dekeyser, P. Bouthemy, and P. Pérez, "Spatio-temporal Wiener filtering of image sequences using a parametric motion model," in *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2000*, Vancouver, Canada, Sept. 2000, pp. 208–211.

[19] S. H. Lee and M. G. Kang, "Spatio-temporal video filtering algorithm based on 3-D anisotropic diffusion equation," in *Proc. of IEEE Int. Conf. on Image Processing, ICIP'1998*, vol. 3(2), Chicago, IL, Oct. 1998, pp. 447–450.

[20] G. Motta, E. Ordentlich, I. Ramirez, G. Seroussi, and M. Weinberger, "The dude framework for continuous tone image denoising," in *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2005*, vol. 3, 2005, pp. 345–348.

[21] S. Awate and R. Whitacker, "Unsupervised, information theoric, adaptive image filtering for image restoration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 364–376, Mar. 2006.

[22] A. Buades, B. Coll, and J. Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, Feb. 2008.

[23] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Proc. of 7th IEEE Int. Conf. on Computer Vision, ICCV'99*, Kerkyra, Greece, Sept. 1999, pp. 1033–1038.

[24] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based inpainting," *IEEE Trans. on Image Processing*, vol. 13, no. 9, pp. 1200–1212, Sept. 2004.

[25] Y. Wexler, E. Shechtman, and M. Irani, "Space-time video completion," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'2004*, vol. 1, Washington, 2004, pp. 120–127.

[26] D. Zhang and Z. Wang, "Image information restoration based on long-range correlation," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 5, pp. 331–341, May 2002.

[27] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *Signal Processing Letters*, vol. 12, no. 12, pp. 839–842, Dec. 2005.

[28] D. Rusanovskyy, K. Dabov, and K. Egiazarian, "Moving-window varying size 3D transform-based video denoising," in *Proc. of 2nd International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM'06*, Jan. 2006.

[29] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionary," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[30] M. Aharon, M. Elad, and A. Bruckstein, "The K-SVD: an algorithm for designing of overcomplete dictionaries for spares representations," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[31] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representation for image and video restoration," *SIAM J. Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.

[32] J. Boulanger, C. Kervrann, and P. Bouthemy, "Space-time adaptation for patch-based image sequence restoration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1096–1102, June 2007.

[33] D. Uttenweiler, C. Weber, B. Jähne, R. Fink, and H. Scharr, "Spatio-temporal anisotropic diffusion filtering to improve signal to noise ratios and object restoration in fluoresence microscopic image sequences," *Journal of Biomedical Optics*, vol. 8, no. 1, pp. 40–47, Jan. 2003.

[34] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser, "Multiframe SURElet denoising of timelapse fluorescence microscopy images," in *Proc. of IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, ISBI'2008*, Paris, France, May 2008, pp. 149–152.

[35] O. Lepski, "Asymptotically minimax adaptive estimation 1: upper bounds," *Theory of Probability and Applications*, vol. 36, no. 4, pp. 654–659, 1991.

[36] O. Lepski, E. Mammen, and V. Spokoiny, "Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors," *Annals of Statistics*, vol. 25, no. 3, pp. 929–947, 1997.

[37] C. Kervrann and A. Trubuil, "An adaptive window approach for poisson noise reduction and structure preserving in confocal microscopy," in *Proc. of IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, ISBI'2004*, Arlington, VA, Apr. 2004, pp. 788–791.

[38] F. Murtagh, J. Starck, and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support," *Astronomy and Astrophysics*, vol. 112, pp. 197–189, 1995.

[39] J.-L. Starck and F. Murtagh, "Automatic noise estimation from the multiresolution support," *Publications of the Astronomical Society of the Pacific*, vol. 110, pp. 193–199, 1998.

[40] J. Boulanger, C. Kervrann, and P. Bouthemy, *Biophotonics for Life Sciences and Medicine*. Fontis Media SA, Lausanne, Switzerland, 2006, ch. An adaptive statistical method for 4D-fluorescence image sequence denoising with spatio-temporal discontinuities preserving, pp. 97–113.

[41] T. Gasser, L. Sroka, and C. Jennen-Steinmetz, "Residual variance and residual pattern in nonlinear regression," *Biometrika*, vol. 73, pp. 625–633, 1986.

[42] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *Multiscale Modeling & Simulation*, vol. 6, no. 2, pp. 595–630, 2007.

[43] N. Azzabou, N. Paragios, and F. Guichard, "Variable bandwidth image denoising using image-based noise models," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'2007*, minneapolis, June 2007, pp. 1–7.

[44] T. Brox and D. Cremers, "Iterated nonlocal means for texture restoration," in *Proc. of Scale Space and Variational Methods in Computer Vision*, A. M. F. Sgallari and N. Paragios, Eds., vol. 4485. Ischia, Italy: Springer, May 2007, pp. 13–24.

[45] S. Kindermann, S. Osher, and P. Jones, "Deblurring and denoising of images by nonlocal functionals," *SIAM J. Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1091–1115, 2005.

[46] A. B. Tsybakov, *Introduction à l'estimation non-paramétrique*, ser. Mathématiques & Applications. Springer, 2003, no. 41.

[47] V. Spokoiny, *Local parametric methods in nonparametric regression*. Springer, 2006.

[48] L. Stankovic, "Performance analysis of the adaptive algorithm for bias-to-variance tradeoff," *IEEE Trans. on Signal Processing*, vol. 52, no. 5, pp. 1228–1234, May 2004.

[49] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. on Image Processing*, vol. 15, no. 10, pp. 2866–2878, 2006.

[50] D. Agard, "Optical sectioning microscopy: cellular architecture in three dimensions," *Annual Review of Biophysics and Bioengineering*, vol. 13, pp. 191–219, June 1984.

[51] J.-B. Sibarita, *Microscopy Techniques*, ser. Advances in Biochemical Engineering/Biotechnology. Berlin / Heidelberg: Springer Verlag, 2005, vol. 95, ch. Deconvolution Microscopy, pp. 201–244.

[52] J.-B. Sibarita, H. Magnin, and J. R. De Mey, "Ultra-fast 4D microscopy and high throughput distributed deconvolution," in *Proc. of IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, ISBI'2002*, Washington, DC, June 2002, pp. 769–772.

[53] S. Hell, "Far-field optical nanoscopy," *Science*, vol. 316, no. 5828, pp. 1153–1158, May 2007.

[54] C. Kervrann and J. Boulanger, "Local adaptivity to variable smoothness for exemplar-based image denoising and representation," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 45–69, Aug. 2008.

Appendix

The estimator $\widehat{u}_{i,n}$ is usually decomposed as

$$\widehat{u}_n(x) = u(x) + b_n(x) + \nu_n(x), \tag{36}$$

where the stochastic term $\nu(x)$ is assumed to be a Gaussian random variable with variance $\upsilon_n^2(x)$. The following inequality holds with a high probability if κ is large enough

$$\exists \kappa \in]0, \infty[: \ |\widehat{u}(x) - u(x)| \leq |b_n(x)| + \kappa \upsilon_n(x). \tag{37}$$

Furthermore, as soon as $|b_n(x)| \leq \gamma \upsilon_n(x)$, we have

$$\exists \kappa \in]0, \infty[: \ |\widehat{u}(x) - u(x)| \leq \gamma \upsilon_n(x) + \kappa \upsilon_n(x). \tag{38}$$

The optimal bandwidth $h^*(x)$ is such that

$$h^*(x) = \sup_{h_n(x) \in \mathcal{H}} \{h_n(x) : |\widehat{u}_n(x) - u(x)| \leq (\gamma + \kappa)\upsilon_n(x)\}. \tag{39}$$

This new inequality is weaker but still explicitly depends on the unknown function. In order to define a practical stopping rule, we consider the following pairwise comparison of successive estimates:

$$\begin{aligned} \widehat{u}_n(x) - \widehat{u}_{n'}(x) &= (u(x) + b_n(x) + \nu_n(x)) \\ &\quad -(u(x) + b_{n'}(x) + \nu_{n'}(x)) \\ &= b_n(x) - b_{n'}(x) + \nu_n(x) - \nu_{n'}(x). \end{aligned}$$

It follows that the random variable $\widehat{u}_n(x) - \widehat{u}_{n'}(x)$ is Gaussian distributed with mean $b_n(x) - b_{n'}(x)$ and variance $\text{Var}[\widehat{u}_n(x) - \widehat{u}_{n'}(x)]$. Moreover, we can prove that $\text{Var}[\widehat{u}_n(x) - \widehat{u}_{n'}(x)] \leq \upsilon_{n'}^2(x)$ knowing that two estimates $\widehat{u}_n(x)$ and $\widehat{u}_{n'}(x)$ and not independent. Hence,

$$\begin{aligned} |\widehat{u}_{n'}(x) - \widehat{u}_n(x)| &\leq |b_{n'}(x) - b_n(x) + \kappa(\text{Var}[\widehat{u}_{n'}(x) - \widehat{u}_n(x)])^{1/2}| \\ &\leq |b_{n'}(x) - b_n(x)| + \kappa \upsilon_{n'}(x) \\ &\leq |b_{n'}(x)| + |b_n(x)| + \kappa \upsilon_{n'}(x)\,. \end{aligned}$$

While $|b_n(x)| \leq \gamma \upsilon_n(x)$, we have

$$\begin{aligned} |\widehat{u}_{n'}(x) - \widehat{u}_n(x)| &\leq \gamma \upsilon_{n'}(x) + \gamma \upsilon_n(x) + \kappa \upsilon_{n'}(x) \\ &\leq (2\gamma + \kappa)\upsilon_{n'}(x). \end{aligned}$$

because $\upsilon_{n'}(x) > \upsilon_n(x)$ for all $n' < n$. From this inequality, we can get the following rule:

$$h^*(x) = \sup_{h_n(x) \in \mathcal{H}} \{n' < n : |\widehat{u}_n(x) - \widehat{u}_{n'}(x)| \leq (2\gamma + \kappa)\upsilon_{n'}(x)\}. \tag{40}$$

This practical rule can be computed since it depends on successive estimates and variances. Finally, we note that $\varrho = 2\gamma + \kappa$ and we get

$$h^*(x) = \sup_{h_n(x) \in \mathcal{H}} \{n' < n \ : \ |\widehat{u}_n(x) - \widehat{u}_{n'}(x)| \leq \varrho \upsilon_{n'}(x)\}. \tag{41}$$

We can further prove that the risk of the estimator (see [54]) is proportional to the risk of the optimal estimator:

$$\mathbb{E}\left[(\widehat{u}(x) - u(x))^2\right] \mathbf{1}_{h^*(x) \leq h(x)} \leq \left(\frac{2\gamma + \kappa}{\sqrt{1+\gamma^2}} + 1\right)^2 \mathcal{R}(u^*(x), u(x)).$$

Centre de recherche INRIA Rennes – Bretagne Atlantique
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399